DOCUMENT RESUME

ED 035 018 CG 004 689

AUTHOR	Flaugher, Ronald L.; And Others
TITLE	Prediction of Job Performance for Negro and White Medical Technicians; Ethnic Group Membership As A Moderator of Supervisor's Ratings.
INSTITUTION	Civil Service Commission, Washington, D.C.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Ford Foundation, New York, N.Y.
REPORT NO	PR-69-5
PUB DATE	Apr 69
NOTE	16p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.90
DESCRIPTORS	Caucasians, Evaluation Methods, *Evaluative Thinking, Job Analysis, *Jobs, Job Skills, Medical Laboratory Assistants, Negro Employment, *Negroes, *Performance Criteria, Prediction, Predictive Measurement, Racial Discrimination, Rating Scales, *Supervisors, Tests

ABSTRACT

A total of 168 Negro and 296 white Medical Technicians participated in this study. A nine-scale performance rating form was developed and administered. Every technician was rated by at least two supervisors. A test of job knowledge was also given to all subjects. Among the correlations found were: (1) Negro workers received distinctly higher ratings on all nine scales from supervisors of their own ethnic group than from white supervisors, (2) Negro raters tended to give higher ratings in general than white raters when rating the same individuals, but the effect was more marked for Negro subordinates than for whites; (3) different relationships exist between scores on the Job Knowledge Test and the criterion of Job Knowledge ratings, depending upon the interaction of the race of the rater and that of the person being rated. Basically, the measures showed that the race of both the person being rated and the person doing the rating has a noteworthy influence on the evaluation received. (KJ)

ED035018

CG004 689

PR-69-5

PREDICTION OF JOB PERFORMANCE FOR NEGRO AND WHITE MEDICAL TECHNICIANS

Ethnic Group Membership as a Moderator of Supervisor's Ratings

Ronald L. Flaugher, Joel T. Campbell,
and Lewis W. Pike

April 1969

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

ED035018

Personnel

Management Committee

Albert P. Maslow, Chief, Personnel Measurement Research and Development Center, U.S. Civil Service Commission

William W. Turnbull, Executive Vice-President, Educational Testing Service

Advisory Committee

John K. Hemphill, Far West Laboratory for Educational Research and Development, Chairman

Marvin D. Dunnette, University of Minnesota

Robert M. Guion, Bowling Green State University

S.O. Roberts, Fisk University

Project Staff

From U.S. Civil Service Commission

William A. Gorham, Program Director

From Educational Testing Service

Joel T. Campbell, Senior Research Psychologist, Principal Investigator

Ronald L. Flaugher, Research Psychologist

Lewis W. Pike, Research Psychologist

Donald A. Rock, Research Psychologist

Franklin R. Evans, Associate Research Psychologist

Mrs. Lila Norris, Associate Research Psychologist

Mrs. Margaret Mahoney, Research Assistant

Mrs. Virginia Rau, Administrative Assistant

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Background[1]

In any study of test validity, the quality of the criterion measures is of critical importance. This fact is not newly discovered nor is it difficult to comprehend, but for a number of reasons it continues to receive far less acknowledgement and attention than it should. The criterion is particularly important when the purpose of the test validation is that of determining the possible existence and the nature of ethnic or cultural bias. The criterion problem is easily overlooked in the concentration upon that which is being validated, and upon the resulting personal and social implications of the findings, but for the very reason that the findings are given such attention, even more thought should be given to the development of high quality criterion measures. It should be clear, for example, that if the chosen criterion is itself infused with the type of bias or unfairness that is being investigated in the predictor tests, the most elaborate of statistical manipulations is unlikely to detect its presence in the predictor.

The present report documents findings derived from an attempt at criterion development for a larger study on fairness in selection testing. Sufficiently large samples were obtained to permit the intensive study of variables which might moderate the criterion scores and thus determine the outcome of the validation attempt. In the present case, the criterion of supervisor's performance ratings was examined for evidence of whether the ethnic group membership of both the rater and the ratee influenced the appraisal an individual received.

---

[1]This study was funded by the Ford Foundation and conducted jointly by the Educational Testing Service and the United States Civil Service Commission.

A total of 168 Negro and 296 white Medical Technicians, employed at 36 U.S. Veterans' Administration hospitals throughout the country, voluntarily participated in this study. As part of the attempt to derive criterion measures against which the validity and fairness of a number of aptitude measures could be judged, a nine-scale performance rating form was developed and administered. Every technician was rated by at least two supervisors. His average rating was used in most subsequent data analysis. The scales on the rating form consisted of descriptions of job performance derived from interviews and field observations of job incumbents. In addition, a 63-item test of technical information about medical laboratory work was developed to serve as a separate criterion measure, and was administered to all incumbents. The rating scales and the test of job knowledge are described in greater detail elsewhere (Project Report 68-4).

Analyses

Mean Scores and Ratings

The means and standard deviations on the nine rating scales and the Job Knowledge Test are shown in Table 1, for technicians grouped by race. White incumbents obtained higher means on all rating scales except "Flexibility," and the difference on that scale was one of the two smallest. Using a two-tailed test, the Wilcoxon matched-pairs signed-ranks test indicates that over the nine rating scales collectively, the difference between ratings of Negro and white incumbents is significant at the .01 level. The difference between the two Job Knowledge Test means was shown by a *t*-test also to be significant at the .01 level, with a *t* of 4.40.

The data were further divided by the ethnic identification of the supervisor doing the rating. Means and standard deviations for these groups

are shown in Table 2. Note that for both Negro and white ratees, the total number of subjects indicated in Table 2 exceeds the number shown in Table 1. This is because each technician rated by supervisors of both ethnic groups was counted in both appropriate columns in Table 2.

The data in Table 2 show a marked influence on the ratings of Negro incumbents depending on whether the rater was of the same race. This was not true of white incumbents. Negro incumbents received distinctly higher ratings on all nine scales from supervisors of their own ethnic group than from white supervisors. In most instances the difference between means was about half of one standard deviation. In this set of nine differences in mean rating, the Wilcoxon test again indicated an effect significant at the .01 level. However, the difference between Job Knowledge Test means for the same two groups was slight, and the t value was only .93. The tendency for ratings of Negro incumbents to differ systematically according to race of rater was not observed for white technicians. The latter received higher mean ratings from white supervisors on four of the scales, and from Negro supervisors on the remaining five. All of the differences were slight, and the Wilcoxon test showed no significant differences between the two sets of ratings. The difference between Job Knowledge Test means was again slight, yielding an observed t of only 1.17.

A possible confounding factor in the latter comparisons is that Negroes rated by whites are not necessarily equivalent in ability to those rated by Negroes. The same is true of the two white ratee groups. If it happens that Negro incumbents working under Negro supervisors are in fact more competent on the average than those working under white supervisors, then their higher ratings could be quite accurate. Fortunately, each incumbent in this

study had been rated by at least two supervisors, and the possible confounding could be controlled by considering only those rated by at least one member of both ethnic groups. Thirty Negroes and thirty-two whites were found to have been rated by supervisors of both races.

The results from these smaller groups are not reproduced here, but were remarkably similar to those of the previous analysis for Negro technicians, who again received higher ratings from Negro supervisors than from white supervisors on every scale. White technicians received higher ratings from their Negro supervisors on seven of the scales, although they did so on only four of the nine scales in the previous analysis. Thus, Negro raters tended to give higher ratings in general than did white raters when rating the same individuals, but the effect was more marked for Negro subordinates than for whites.

Correlations

Again using the combined ratings on all incumbents, each of the four groups of supervisors' ratings were then correlated with the technicians' Job Knowledge Test scores. The results appear in Table 3.

Most noticeable about those results are the generally low correlates of the objective test with the ratings given to white technicians by Negro supervisors. Those factors measured by the Job Knowledge Test appear to have a sizable influence on white supervisor's ratings of both Negro and white technicians, and on Negro supervisor's ratings of Negroes. Negro supervisor's ratings of white technicians, however, appear to be essentially unrelated to factors measured by the Job Knowledge Test. The pattern of correlations across rater-ratee ethnic combinations just described is particularly evident for Job Knowledge ratings, where the basis of rating

should most nearly coincide with what is measured by the Job Knowledge Test.

Regression Lines

In order to illustrate the relationships more clearly, a hypothetical situation may be considered, in which it could be asked what the results would be if the Job Knowledge Test were used as a prediction device, and the Job Knowledge rating scale were employed as the criterion to be predicted. Regression lines for each of the rater-ratee groups can be drawn, incorporating the means and variances of both predictor and criterion, as well as the correlation between them. Figure 1 shows that rather dramatically different relationships exist between scores on the Job Knowledge Test and the criterion of Job Knowledge ratings, depending upon the interaction of the race of the rater and that of the person being rated.

The possibility was mentioned earlier of a confounding effect due to actual differences between, say, Negro incumbents working under Negro supervisors and those working under whites. The regression lines shown in Figure 2 are based on data for subjects rated by both Negro and white supervisors, thus eliminating this confounding effect. As is evident in Figure 2, the results are essentially the same as those obtained for all subjects.

Discussion

From the hypothetical use of a job knowledge test as a predictor of ratings described above, it is clear that very different conclusions would be reached about the usefulness of the predictor test depending upon the ethnic composition of the rater and ratee groups. For three of the four possible combinations the test would have some useful validity, but for the fourth, it would have none.

An informal attempt was made to account for the differences in ratings associated with the four ethnic combinations of rater and ratee found in the present study. As part of the larger study, all technicians had taken a battery of aptitude tests. This permitted an examination of the correlates of the rating scales with tests of known aptitude factors. The complete data are available elsewhere, but a partial repetition here will serve to clarify the findings described above.

For the data based on ratings of white technicians by Negro supervisors, the distinctive part of the pattern of correlations with the rating scales was the appearance of rather sizable negative correlations between the ratings and the Hidden Figures Test, as shown in Table 4. Thus, while the correlations between the ratings and the Job Knowledge Test were comparatively low for this group, the correlations between the ratings and Hidden Figures were high, and in a negative direction.

Fortunately, the Hidden Figures Test is well-studied, and the unusual pattern of correlations found between the ratings and scores on that test suggests an explanation of the influences operating in Negro supervisors' ratings of whites. The Hidden Figures Test has been used frequently in studies of cognitive style (Witkin, et al., 1962), and is known as a measure of "Field independence versus field dependence." Persons who do poorly on this task are more distracted by the embedding context than those who do well, and they tend to be more dependent in their interpersonal relations.

A very tentative conclusion which might be derived, then, is that the particularly independent individual, if white, is seen unfavorably by a Negro supervisor. In other ethnic combinations of supervisor-subordinate

relations, supervisory judgments appear to be rather firmly related to the technician's knowledge of his job. For the Negro-rating-white combination, however, the personality of the subordinate would appear to play a more significant role than job knowledge in determining his performance ratings.

Care must be taken in attempting to generalize from the specific results of the present study to any situation other than that in which they were obtained. A generalization from these results which does seem appropriate, however, is that close attention must be paid to the nature of the criterion data employed in any prediction study; unperceived combinations of influences operating upon the criterion measures may determine the results of any analysis fully as much as the content or structure of the predictive measure being evaluated.

Summary

The interrelationships of several criterion measures are reported, showing that the race of both the person being rated and the person doing the rating has a noteworthy influence on the evaluation received. Each of the four possible combinations produced a different structure of correlates. An argument was presented in favor of attending closely to the criterion against which any predictor test is being evaluated.

References

Pike, L.W. Prediction of job performance for Negro and white medical technicians: Development of the instrumentation. Project Report 69-4. Princeton, New Jersey: Educational Testing Service, 1969.

Witkin, H.A., Dyk, R.B., Faterson, H.F. Goodenough, D.R., and Karp, S.A. Psychological Differentiation. New York: Wiley, 1962.

Table 1

Supervisory Ratings and Job Knowledge Test Scores

of Negro and White Medical Technicians

Criterion Measures	Ethnic Group of Ratee			
	Negro (N = 165 - 166)		White (N = 283 - 297)	
	Mean	S.D.	Mean	S.D.
Rating Scales				
Flexibility	5.5	1.8	5.4	1.8
Organization	5.7	1.8	6.0	1.8
Interest	5.6	1.8	5.9	1.7
Learning Ability	5.8	1.9	6.0	1.8
Job Knowledge	5.2	1.9	5.3	1.7
Technique	5.9	1.7	6.0	1.7
Low Need for Supervision	5.7	2.0	6.1	1.9
Communication	5.5	1.8	5.8	1.8
Overall	5.7	1.9	5.9	2.0
Job Knowledge Test	31.4	10.4	35.7	9.4

Table 2

Supervisory Ratings and Job Knowledge Test Scores

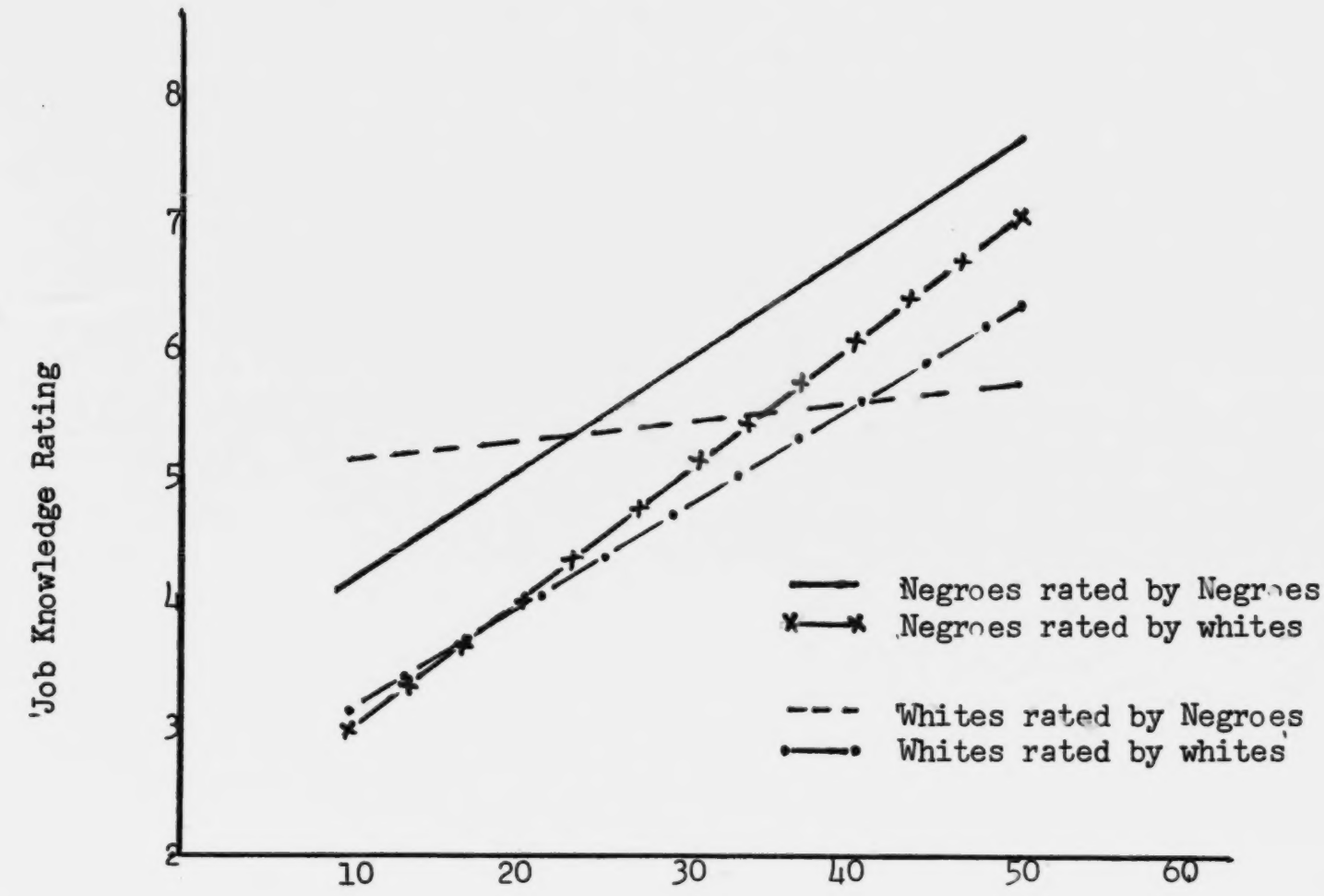
Grouped by Rater-Ratee Ethnic Combination

Criterion Measures	Negro rated by Negro (N = 43-50)		Negro rated by white (N = 146-148)		White rated by Negro (N = 38-41)		White rated by white (N = 272-276)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rating Scales								
Flexibility	6.2	2.0	5.3	1.9	5.5	2.2	5.4	1.8
Organization	6.3	1.9	5.6	1.9	6.1	2.2	6.0	1.8
Interest	6.1	2.0	5.4	1.8	6.2	1.9	5.8	1.8
Learning Ability	6.6	2.0	5.6	2.0	5.8	2.1	6.0	1.8
Job Knowledge	6.2	2.0	5.1	1.9	5.5	1.8	5.3	1.8
Technique	6.5	1.7	5.8	1.8	5.8	1.7	6.0	1.8
Low Need for Supervision	6.5	2.0	5.6	2.0	5.9	2.2	6.1	1.9
Communication	6.3	2.0	5.3	1.8	6.0	1.7	5.7	1.8
Overall	6.3	2.0	5.6	2.0	5.8	2.2	5.9	2.0
Job Knowledge Test	32.6	10.8	31.0	10.6	34.2	11.2	36.0	9.0

Table 3

Correlations of Supervisory Ratings
with the Test of Job Knowledge

Rating Scale	Negro rated by Negro (N = 34-40)	Negro rated by white (N = 136-138)	White rated by Negro (N = 35-39)	White rated by white (N = 264-268)
Flexibility	.14	.50	.18	.32
Organization	.35	.41	.24	.27
Interest	.41	.35	.07	.25
Learning Ability	.59	.53	.28	.36
Job Knowledge	.50	.56	.09	.39
Technique	.41	.46	.18	.25
Low Need for Supervision	.41	.40	.16	.29
Communication	.54	.47	.08	.30
Overall	.41	.42	.09	.28



Job Knowledge Test Scores

Fig. 1. Prediction of Job Knowledge Rating from Job Knowledge Test score, using data from all technicians.

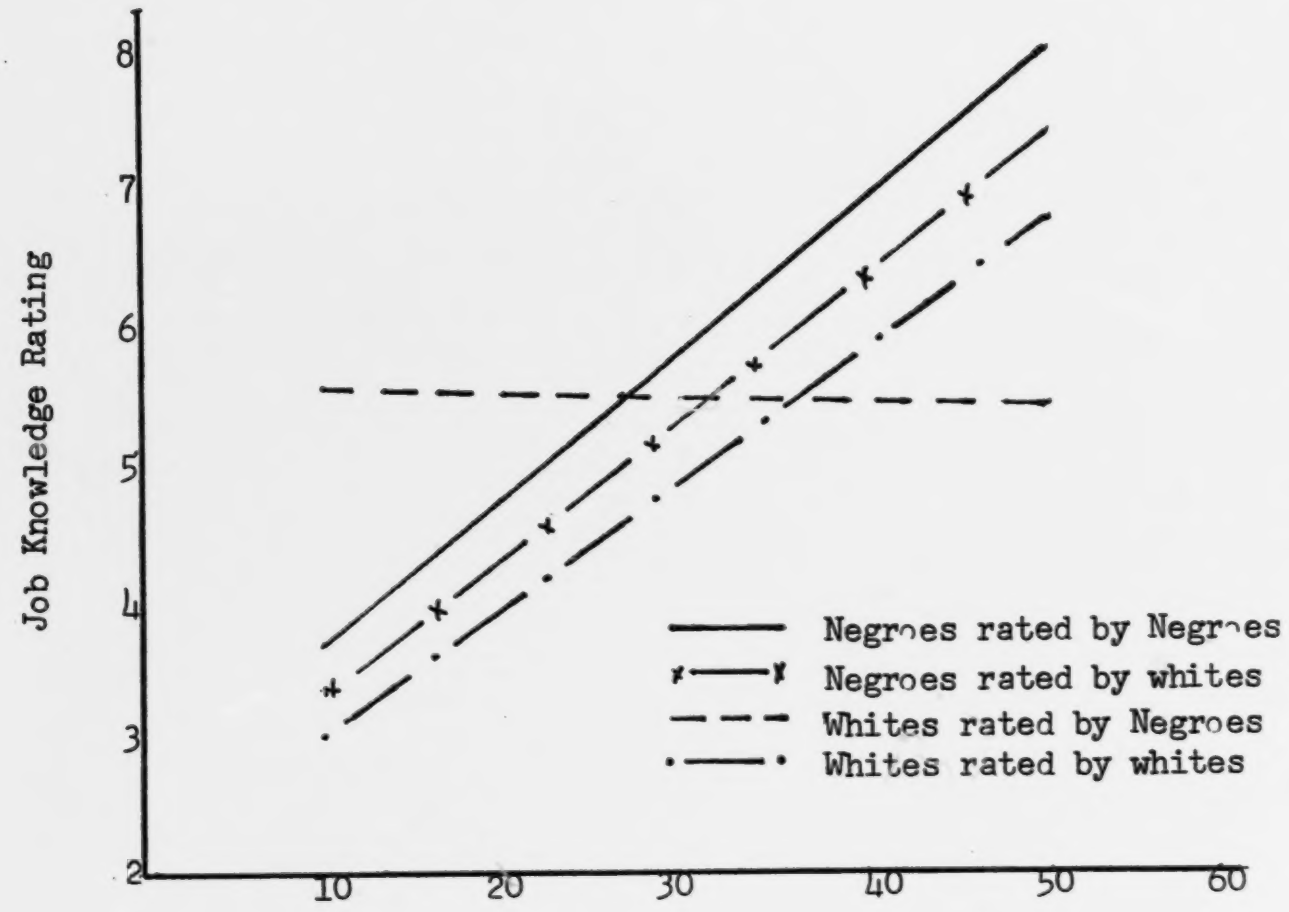


Fig. 2. Prediction of Job Knowledge Rating from Job Knowledge Test score, using data from technicians rated by both Negro and white supervisors.

Table 4

Correlations of Supervisory Ratings With

Technicians' Scores on the Hidden Figures Test

Rating Scale	Negro rated by Negro (N = 34-40)	Negro rated by white (N = 136-138)	White rated by Negro (N = 35-39)	White rated by white (N = 264-268)
Flexibility	-.10	.04	-.30	.20
Organization	.11	.00	-.32	.07
Interest	.05	.03	-.29	.06
Learning Ability	-.12	-.01	.01	.13
Job Knowledge	-.01	.06	.05	-.00
Technique	-.14	.03	-.11	.08
Low Need for Supervision	-.12	.01	-.28	.06
Communication	-.08	-.04	-.08	.08
Overall	-.07	-.01	-.17	.07